

Pr. BENSOLTANE FOUAD

I- Série statistique à caractère discret :

1/ Tableau d'effectifs, d'effectifs cumulés, de fréquences et de fréquences cumulées:

Après avoir corrigé un contrôle de maths d'une classe de 3AC le prof a présenté les données suivantes : 10 ; 14 ; 11 ; 9 ; 12 ; 10 ; 14 ; 10 ; 17 ; 9
10 ; 11 ; 12 ; 10 ; 14 ; 9 ; 10 ; 9 ; 10 ; 11

Valeurs de caractère (les notes)	9	10	11	12	14	17
Effectifs	4	7	3	2	3	1
Effectifs cumulés	4	11	14	16	19	20
Fréquences = $\frac{\text{Effectif}}{\text{Effectif total}}$	$\frac{4}{20}$	$\frac{7}{20}$	$\frac{3}{20}$	$\frac{2}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
Fréquences cumulées	$\frac{4}{20}$	$\frac{11}{20}$	$\frac{14}{20}$	$\frac{16}{20}$	$\frac{19}{20}$	1

2/ Effectif total:

a) **Définition:** L'effectif total d'une série statistique, c'est la somme des effectifs, noté : N

b) **Exemple:** On a d'après l'exemple ci-dessus : $N = 4 + 7 + 3 + 2 + 3 + 1 = 20$

Donc l'effectif total de cette série statistique est : $N = 20$.

3/ Les caractéristiques d'une série statistique à caractère discret:

a) Le mode :

a₁) **Définition:** Le mode d'une série statistique, c'est la valeur du caractère qui a le plus grand effectif.

a₂) **Exemple:** Dans l'exemple ci-dessus on a :

Le plus grand effectif est 7, associé à la valeur de caractère 10.

Donc : le mode de cette série statistique est : 10.

*/ **Remarque importante :** Une série statistique peut avoir plus d'un mode.

b) La valeur moyenne d'une série statistique :

b₁) **Définition:** La valeur moyenne d'une série statistique, c'est le quotient de la somme des produits des effectifs par leurs valeurs associées par l'effectif total, notée: m

b₂) **Exemple:** Dans l'exemple ci-dessus on a :

$$m = \frac{4 \times 9 + 7 \times 10 + 3 \times 11 + 2 \times 12 + 3 \times 14 + 1 \times 17}{20}$$

$$\text{C'est-à-dire : } m = \frac{36 + 70 + 33 + 24 + 42 + 17}{20} = \frac{222}{20} = 11,1.$$

D'où : la valeur moyenne de cette série statistique est : 11,1.

c) La médiane d'une série statistique :

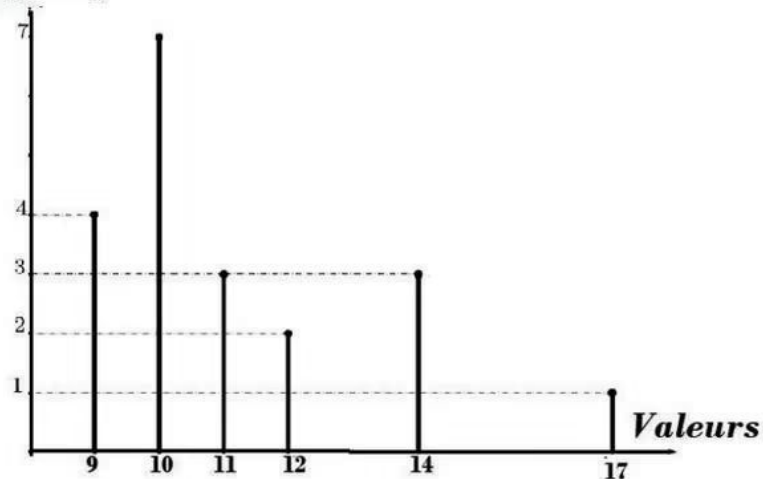
c₁) Définition : La médiane d'une série statistique c'est la valeur associée au plus petit effectif cumulé, supérieur ou égal à la moitié de l'effectif total.

c₂) Exemple : Dans l'exemple ci-dessus on a : $\frac{N}{2} = \frac{20}{2} = 10$.

Le plus petit effectif cumulé supérieur ou égal à 10 est 11 qui est associé à la valeur 10.

Donc : la médiane de cette série statistique est : 10.

4/ Le diagramme en bâton : Effectifs



II - Série statistique à caractère continu :

1/ Tableau d'effectifs, d'effectifs cumulés, de fréquences et de fréquences cumulées:

Le tableau suivant représente les tailles des ouvriers d'une usine, données par groupe.

Les classes	$120 \leq t < 130$	$130 \leq t < 140$	$140 \leq t < 150$	$150 \leq t < 160$
Effectifs	9	18	12	11
Effectifs cumulés	9	27	39	50
Fréquences = $\frac{\text{Effectif}}{\text{Effectif total}}$	$\frac{9}{50}$	$\frac{18}{50}$	$\frac{12}{50}$	$\frac{11}{50}$
Fréquences cumulées	$\frac{9}{50}$	$\frac{27}{50}$	$\frac{39}{50}$	1

2/ **Effectif total:** On a : $N = 9 + 18 + 12 + 11 = 50$

Donc l'effectif total de cette série statistique est : $N = 50$.

3/ Les caractéristiques d'une série statistique à caractère continu :

a) Le mode :

a₁) Définition: Le mode d'une série statistique à caractère continu , c'est la classe qui a le plus grand effectif.

a₂) Exemple: Dans l'exemple ci-dessus on a :

Le plus grand effectif est 18, associé à la classe $130 \leq t < 140$

Donc : le mode de cette série statistique c'est la classe : $130 \leq t < 140$.

b) Le centre d'une classe :

b₁) Définition: Le centre d'une classe $a \leq x < b$ est $\frac{a+b}{2}$

b₂) Exemple: On considère le tableau suivant :

Centres des classes	125	135	145	155
Les classes	$120 \leq t < 130$	$130 \leq t < 140$	$140 \leq t < 150$	$150 \leq t < 160$
Effectifs	9	18	12	11

c) La moyenne pondérée d'une série statistique :

c₁) Définition: La moyenne pondérée d'une série statistique à caractère continu, c'est le quotient de la somme des produits des effectifs par leurs le centre de classe associées par l'effectif total, notée : m

c₂) Exemple: Dans l'exemple ci-dessus on a :

$$m = \frac{125 \times 9 + 135 \times 18 + 145 \times 12 + 155 \times 11}{50}$$

$$\text{C'est-à-dire : } m = \frac{7000}{50} = 140.$$

D'où : la valeur moyenne de cette série statistique est : 140

d) La médiane :

d₁) Définition :

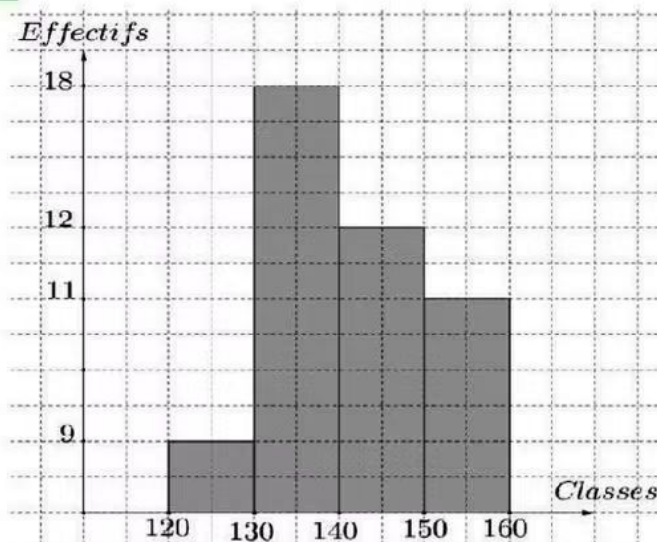
La médiane d'une série statistique à caractère continu se trouve dans la classe associée au plus petit effectif cumulé, supérieur ou égal à la moitié de l'effectif total.

d₂) Exemple : Dans l'exemple ci-dessus on a : $\frac{N}{2} = \frac{50}{2} = 25.$

Le plus petit effectif cumulé supérieur ou égal à 25 est 27, associé à la classe : $130 \leq t < 140$

Donc : la médiane de cette série statistique se trouve dans la classe : $130 \leq t < 140$.

4/ L'histogramme :



Chapitre XIII

Statistiques

I) Statistiques à une variable quantitative :

N : effectif total, $N = n_1 + n_2 + \dots + n_p$.

• Variable continue :

Classe	$[a_0, a_1[$		$[a_{p-1}, a_p[$
Effectifs n_i	n_1		n_p
Fréquence f_i	$\frac{n_1}{N}$		$\frac{n_p}{N}$

Centre de classe $[a_{k-1}, a_k[$ est $x_k = \frac{a_{k-1} + a_k}{2}$

• Variable discontinue où discrète :

Valeur x_i	x_1	x_2		x_p
Effectifs n_i	n_1	n_2		n_p
Fréquences f_i	$\frac{n_1}{N}$	$\frac{n_2}{N}$		$\frac{n_p}{N}$

- La Fréquence en pourcentage est $\frac{n_i}{N} \times 100$

Paramètres de position :

■ **Moyenne simple:** $\bar{X} = \frac{x_1 + x_2 + x_3 + \dots + x_k}{k} = \frac{1}{k} \sum_{i=1}^k x_i$

- **Moyenne pondérés:** Moyenne de la série (x_i, n_i)

$$\bar{X} = \frac{n_1 \cdot x_1 + n_2 \cdot x_2 + \dots + n_k \cdot x_k}{N} = \frac{1}{N} \sum_{i=1}^k x_i \cdot n_i \text{ et } N = \sum_{i=1}^k n_i$$

Paramètres de dispersion :

■ Variance:

* Variance de la série x : $V_x = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$

* Variance de la série (x_i, n_i) : $V_x = \frac{1}{N} \sum_{i=1}^k n_i \cdot x_i^2 - \bar{x}^2 = \frac{1}{N} \sum_{i=1}^k n_i \cdot (x_i - \bar{x})^2$

- **Ecart - type:** Dans tous les cas on a : $\sigma_x = \sqrt{V_x}$.

Remarque :

- Une série ayant un faible écart-type est une série dont les valeurs sont peu dispersées.

- Une série ayant un écart-type important est une série dont les valeurs sont très dispersées.

■ Diagramme en boîte :

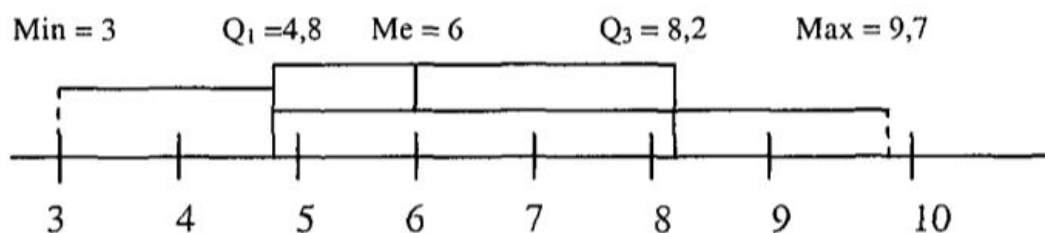
Les deux quartiles Q_1 , Q_3 et la médiane Me d'une série statistique associés aux valeurs extrêmes (minimum et maximum) permettent d'appréhender certaines caractéristiques de la répartition des données. On les représente souvent par un diagramme en boîte.

Exemple :

Voici la liste des températures relevées sous abri à différents moments d'une journée, rangées par ordre croissant et exprimées en degrés Celsius.

3- 3,8 - 4,5 - 5- 5,5 -5,7 - 5,8- 6,2- 7- 7, 3- 8,2 -9-9,2 - 9,5 - 9,7

Voici le diagramme en boîte de cette série .



Ce diagramme est souvent employé pour comparer la répartition de deux séries.

■ L'écart interquartile :

L'écart interquartile relatif d'une série statistique à une variable quantitative est égale au quotient de l'écart interquartile par la médiane de la série. C'est un nombre sans unité qui peut être exprimé en pourcentage.

On utilise l'écart interquartile relatif lorsqu'on veut comparer les dispersions autour de la médiane de deux séries statistiques dont les valeurs respectives ont des ordres de grandeur différents. Plus L'écart interquartile relatif est faible plus la dispersion autour de la médiane est faible.

Soit une série statistique de valeurs (x_1, x_N) de moyenne $\bar{x}$ et l'écart type σ . Soit a, b deux réels et Y la série statistique de valeur $(ax_1 + b, ax_2 + b, \dots, ax_N + b)$. On désigne par $\bar{Y}$ et σ' la moyenne et l'écart type de y alors $\bar{Y} = a\bar{x} + b$ avec $\sigma' = |a|\sigma$

X une série statistique de moyenne $\bar{X}$ d'écart type σ .

On dit que X est une distribution normale ou gaussienne lorsque le polygone des effectifs est tel que environ

- 68 % des effectifs sont situés dans $[\bar{x} - \sigma, \bar{x} + \sigma]$
- 95 % des effectifs sont situés dans $[\bar{x} - 2\sigma, \bar{x} + 2\sigma]$
- 99 % des effectifs sont situés dans $[\bar{x} - 3\sigma, \bar{x} + 3\sigma]$

Dans ce cas, le polygone des effectifs à la forme d'une cloche symétrique par rapport à la moyenne.

Le réel $\frac{\sigma}{\bar{X}}$ est dit l'écart type relatif de X il peut être exprimé en pourcentage

■ Statistiques à deux variables :

- Etude de deux caractères sur une même population 1^{er} type de tableau à double entrée

$y \backslash x$	x_1	x_2		x_p
y_1	n_{11}	n_{12}		n_{1p}
y_2	n_{21}	n_{22}		n_{2p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
y_q	n_{q1}	n_{q2}		n_{qp}

Les x_i et les y_i peuvent être remplacés par des classes pour une variable continue.

- **Nuage des points** : Dans un repère orthogonal $(O, \vec{i}, \vec{j})$ on place les x_i sur l'axe de abscisses et y_i sur l'axe des ordonnées on place les points de coordonnées (x_i, y_i) l'ensemble des points représenté s'appelle **nuage de points**.
- Le point $G(\bar{X}, \bar{Y})$ s'appelle **le point moyen**.

Comment Calculer $\bar{X}$ et $\bar{Y}$? (La notation n_{12} : 1 correspond 1^{ère} ligne et 2 2^{ème} colonne)

$\begin{matrix} x \\ y \end{matrix}$	x_1	x_2		x_p	Distribution marginale de Y
y_1	n_{11}	n_{12}		n_{1p}	$n_{11} + n_{12} + \dots + n_{1p} = N_1$
y_2	n_{21}	n_{22}		n_{2p}	N_2
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
y_q	n_{q1}	n_{q2}		n_{qp}	N_q
Distribution marginale de X	N'_1	N'_2		N'_p	N

$N'_1 = n_{11} + n_{21} + n_{31} + \dots + n_{q1}$ (on fait la somme des effectifs)

N : le nombre total de l'effectif.

$$\bar{X} = \frac{\sum_{i=1}^p x_i N'_i}{N} \text{ et } \bar{Y} = \frac{\sum_{i=1}^q y_i N_i}{N}.$$

- Si on a des classes on utilise leur centre.
- On peut réaliser le même tableau avec les fréquences

Exemple : $f_{2p} = \frac{n_{2p}}{N}.$

2^{ème} type de tableau d'un série à deux caractère :

X	x_1	x_2		x_p
Y	y_1	y_2		y_p

- Le nuage de points : l'ensemble des points $M_i(x_i, y_i)$ placer dans un repère orthogonal.

- Le point moyen $G(\bar{X}, \bar{Y})$ avec $\bar{X} = \frac{\sum_{i=1}^p x_i}{N}$ et $\bar{Y} = \frac{\sum_{i=1}^p y_i}{N}.$

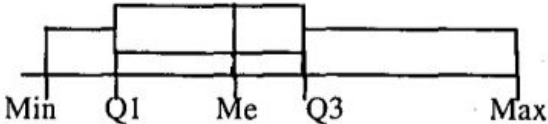
■ Droite d'ajustement affine:

On note P_1 le nuage de points de la 1^{ère} moitié de la série (x, y) et P_2 le nuage de la 2^{ème} moitié de la série. (On divise le tableau statistique en 2 parties quelconques).

- G_1 le point moyen de P_1 et G_2 le point moyen de P_2

La droite (G_1, G_2) est appelée la droite de Mayer ou bien la droite d'ajustement affine du nuage de points représentant la série double (x,y) et elle passe par le point moyen G.

Réflexes :

Situation	Réflexes
Comment déterminer la médiane Me d'une série statistique ?	1) On ordonne les données dans l'ordre croissant : $x_1 \leq x_2 \leq \dots \leq x_n$ 2) Si $n = 2q + 1$ alors $Me = x_{q+1}$ et si $n = 2q$ alors $Me = \frac{1}{2} (x_q + x_{q+1})$
Comment déterminer les quartiles d'une série statistique ?	1) On ordonne les données dans l'ordre croissant : $x_1 \leq x_2 \leq \dots \leq x_n$ 2) $Q_1 = x_i$ où i est le plus petit supérieur ou égal à $\frac{n}{4}$; $Q_3 = x_j$ où j est le plus petit entier supérieur ou égal à $\frac{3n}{4}$.
Comment construire un diagramme en boîte ?	1) On détermine le minimum, le maximum, la médiane Me et les quartiles Q_1 et Q_3 de la série. 2) On construit la boîte le long d'une graduation comme ci-dessous. 
Comment résumer une série statistique ?	Par une mesure de tendance centrale associée à une mesure de dispersion : (moyenne ; écart-type) ou (médiane ; écart interquartile).
Comment retrouver la droite d'ajustement linéaire ? (droite de Mayer)	On divise le tableau en deux parties, on calcule $\bar{x}_1$ et $\bar{y}_1$ de la première partie. Puis $\bar{x}_2$ et $\bar{y}_2$ celle de la 2 ^{ème} partie. $G_1(\bar{x}_1, \bar{y}_1)$ et $G_2(\bar{x}_2, \bar{y}_2)$ alors la droite de Mayer est (G, G_2) puis on écrit son équation $y = ax + b$
Comment estimer la valeur de la variable y pour une valeur X_0 donnée de X, lors de l'ajustement linéaire de y en X ?	Une estimation de la valeur y est $y_0 = aX_0 + b$